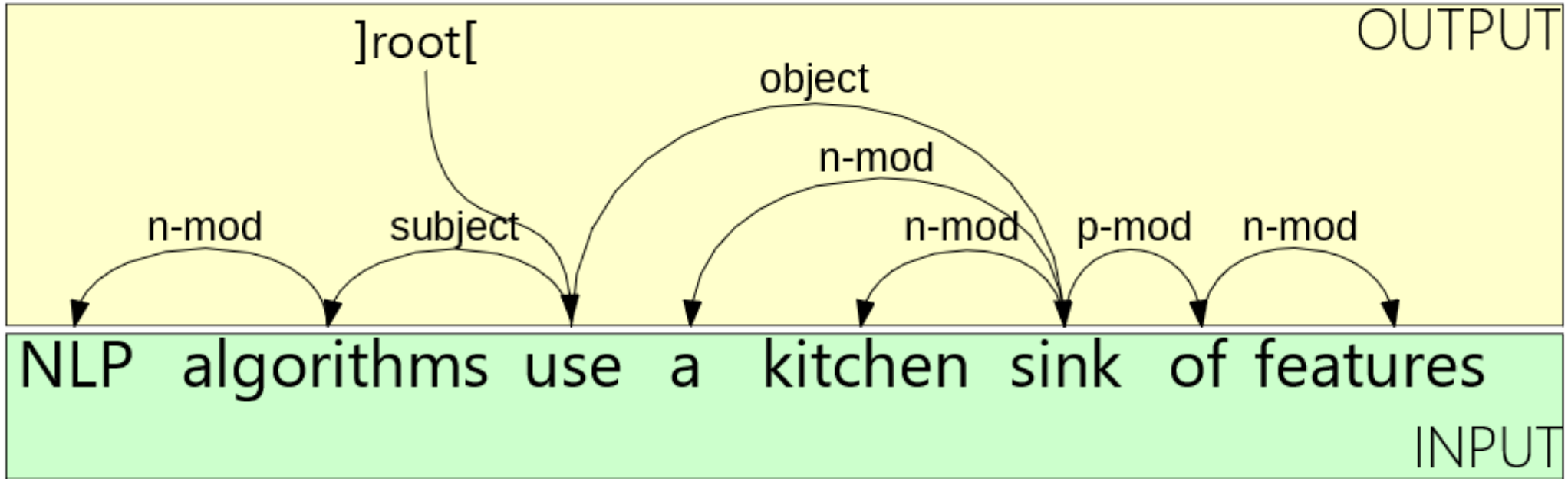
Structured Prediction APML

Omri Abend

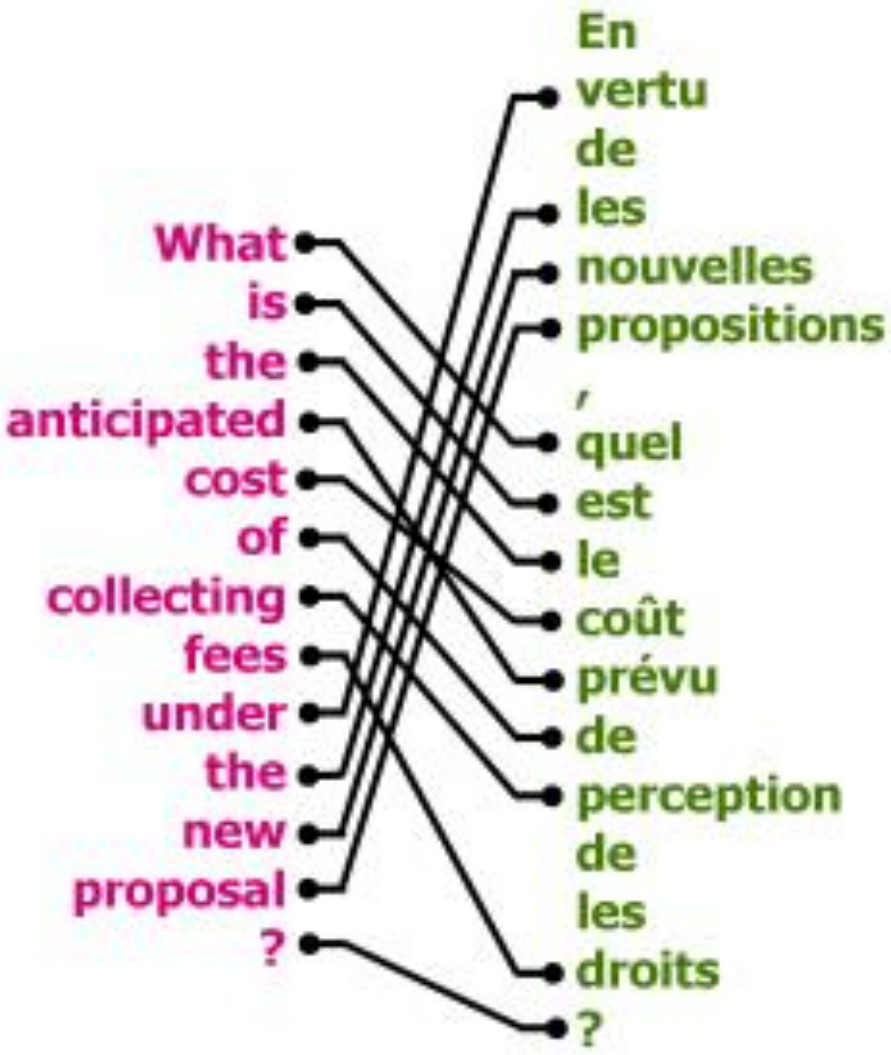
Structured Prediction

- The classic cases of classification are binary or multi-class
- Structured prediction is the case where the label space has some structure
  - The label space is often dependent on the input
- Another way to think about it is that structured prediction involves many inter-dependent predictions

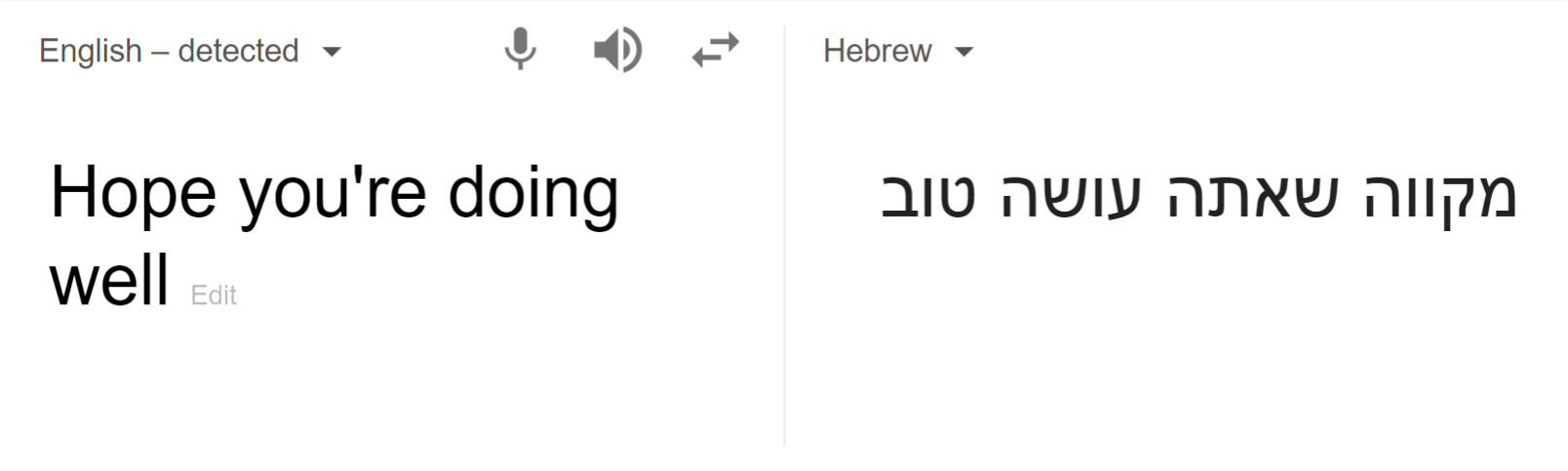
Natural Language Parsing



Bipartite Matching



Machine Translation



Named Entity Recognition (NER)

INPUT: Profits soared at Boeing Co., easily topping forecasts on Wall Street, as their CEO Alan Mulally announced first quarter results.

OUTPUT: Profits soared at [Company Boeing Co.], easily topping forecasts on [Location Wall Street], as their CEO [Person Alan Mulally] announced first quarter results.

Named Entity Recognition (NER)

INPUT:

Profits soared at Boeing Co., easily topping forecasts on Wall Street, as their CEO Alan Mulally announced first quarter results.

OUTPUT:

Profits/NA soared/NA at/NA Boeing/SC Co./CC ,/NA easily/NA topping/NA forecasts/NA on/NA Wall/SL Street/CL ,/NA as/NA their/NA CEO/NA Alan/SP Mulally/CP announced/NA first/NA quarter/NA results/NA ./NA

NA	= No entity
SC	= Start Company
CC	= Continue Company
SL	= Start Location
CL	= Continue Location

...

Part of Speech (POS) Tagging

INPUT:

Profits soared at Boeing Co., easily topping forecasts on Wall Street, as their CEO Alan Mulally announced first quarter results.

OUTPUT:

Profits/N soared/V at/P Boeing/N Co./N ,/, easily/ADV topping/V forecasts/N on/P Wall/N Street/N ,/, as/P their/POSS CEO/N Alan/N Mulally/N announced/V first/ADJ quarter/N results/N ./.

N = Noun
V = Verb
P = Preposition
Adv = Adverb
Adj = Adjective
...

POS Tagging – the Supervised Setup

Training set:

1 Pierre/NNP Vinken/NNP ,/, 61/CD years/NNS old/JJ ,/, will/MD join/VB the/DT board/NN as/IN a/DT nonexecutive/JJ director/NN Nov./NNP 29/CD ./.
2 Mr./NNP Vinken/NNP is/VBZ chairman/NN of/IN Elsevier/NNP N.V./NNP ,/, the/DT Dutch/NNP publishing/VBG group/NN ./.
3 Rudolph/NNP Agnew/NNP ,/, 55/CD years/NNS old/JJ and/CC chairman/NN of/IN Consolidated/NNP Gold/NNP Fields/NNP PLC/NNP ,/, was/VBD named/VBN a/DT nonexecutive/JJ director/NN of/IN this/DT British/JJ industrial/JJ conglomerate/NN ./.

...

38,219 It/PRP is/VBZ also/RB pulling/VBG 20/CD people/NNS out/IN of/IN Puerto/NNP Rico/NNP ,/, who/WP were/VBD helping/VBG Huricane/NNP Hugo/NNP victims/NNS ,/, and/CC sending/VBG them/PRP to/TO San/NNP Francisco/NNP instead/RB ./.

- From the training set, induce a function/algorithm that maps new sentences to their tag sequences.

POS Tagging – the Supervised Setup

Training set:

1 Pierre/NNP Vinken/NNP ,/, 61/CD years/NNS old/JJ ,/, will/MD join/VB the/DT board/NN as/IN a/DT nonexecutive/JJ director/NN Nov./NNP 29/CD ./.
2 Mr./NNP Vinken/NNP is/VBZ chairman/NN of/IN Elsevier/NNP N.V./NNP ,/, the/DT Dutch/NNP publishing/VBG group/NN ./.
3 Rudolph/NNP Agnew/NNP ,/, 55/CD years/NNS old/JJ and/CC chairman/NN of/IN Consolidated/NNP Gold/NNP Fields/NNP PLC/NNP ,/, was/VBD named/VBN a/DT nonexecutive/JJ director/NN of/IN this/DT British/JJ industrial/JJ conglomerate/NN ./.
...
38,219 It/PRP is/VBZ also/RB pulling/VBG 20/CD people/NNS out/IN of/IN Puerto/NNP Rico/NNP ,/, who/WP were/VBD helping/VBG Huricane/NNP Hugo/NNP victims/NNS ,/, and/CC sending/VBG them/PRP to/TO San/NNP Francisco/NNP instead/RB ./.

- From the training set, induce a function/algorithm that maps new sentences to their tag sequences.

Evaluation:

$$accuracy = \frac{\#\ test\ set\ words\ with\ correct\ tag}{\#test\ set\ words}$$

Word tokens as opposed to *word types*. That is of the *word type* dog appears 5 times in the test set, it will be counted 5 times by the accuracy measure

Part of Speech Tags: in more detail

- Wordforms often have more than one possible POS: *back*
  - The *back* door = *Adj*
  - On my *back* = *Noun*
  - Win the voters *back* = Adverb
  - Promised to *back* the bill = Verb

- The POS tagging problem is to determine the (single) POS tag for a particular word token (instance)

Sources of information

- What are the main sources of information for POS tagging?
  1. Knowledge of word probabilities
     - *man* is rarely used as a verb...
  2. Knowledge of neighboring words

Bill	saw	that	man	yesterday
name	**verb (past)**	**det.**	**noun**	**adverb**
verb	*verb*	*det.*	*verb*	*adverb*
verb	*noun*	*conj.*	*verb*	*adverb*

Word-level Classification

- We can do pretty well by classifying each word on its own
  - But we'll struggle with infrequent words
- Orthographic features:
  - Lowercase / uppercase
  - Prefixes / suffixes ('-ed' → verb, '-ly' → adverb, 'un-' → adjective)
  - Non-letter characters (periods → acronyms, only numbers → quantifier)

The four components of a computational account of an (NLP) phenomena:

1. Theory (representation)
2. Statistical model
3. Parameter estimation, a.k.a learning
4. Prediction, a.k.a inference